In the event that a fee is due/overpaid, the Commissioner is hereby authorized to charge/credit any such fees to Deposit Account **02-3964** (Order No. 20206-135).

Respectfully submitted,

Dated: 4-10-02

Leah Sherry
Reg. No. 43,918

OPPENHEIMER WOLFF & DONNELLY LLP
Customer Number 25696
P.O. Box 10356
Palo Alto, CA 94303
Tel: 650/320-4000
Fax: 650/320-4100

CERTIFICATE OF MAILING (37 CFR 1.8(a))

I hereby certify that this paper (along with any specified enclosures) is being deposited with the U.S. Postal Service as first class mail in an envelope addressed to Assistant Commissioner of Patents, Washington, D.C. 20231 on:

Date: 4-10-02

By Tiffany Bell
Tiffany Bell

APPENDIX A
Marked-up Version of Amended Specification Paragraphs

The paragraph starting at page 9, line 8, is amended as follows:

Whereas prior art systems, as described above, assume complete independence between joined columns [is assumed], an embodiment of the invention assumes complete dependence between joined columns in order to determine an estimated selectivity to be passed and used by optimizer 112 of Figure 1. In implementing this embodiment, a minimum single column selectivity from columns A and B is chosen to produce an estimate of the join selectivity. Accordingly, a minimum selectivity value will produce a larger row selectivity (or join selectivity). In applying this method of the present invention to columns A and B, the row selectivity is estimated using the following equation:

S_{dj} = MIN(1/CurUecA, 1/CurUecB) * Xprod

= 1/102 * 80 000

= 784 rows[,]

where the subscript dj denotes a completely dependent join and Xprod represents the cross product of columns A and B. We can rewrite the above equation by applying certain identities. We note that row selectivities for columns A and B are respectively the values:

RowSelA = 1/CurUecA, and

RowSelB = 1/CurUecB

where

RowSelA is the row selectivity for column A, and

RowSelB is the row selectivity for column B.

Applying these identities, we can write

S_{dj} = MIN(RowSelectA, RowSelectB) * Xprod

= 1/102 * 80 000

= 784 rows.

Moreover, note that the above equation is equivalent to choosing a maximum single column current UEC such that the applied equation can be written in an alternative form:

S_{dj} = 1/MAX(CurUecA , CurUecB) * Xprod

= 1/102 * 80 000

= 784 rows

where S_{dj} is the selectivity as defined above.

It has been found that where skew and possible row and UEC reduction can be ignored this estimate provides a much improved estimate of selectivity than one derived assuming complete independence.

Where such conditions are met, the estimated selectivity of 784 rows is much improved from the dramatic underestimate for selectivity of 65 rows obtained using the prior art method.

The paragraph starting at page 12, line 13, is amended as follows:

______In order to obtain, multi-column histogram information, we apply the following formula to each interval shown above:

(XprodT1.A) * (XprodT2.A) / MAX ((CurUECT1.A~~Current UECT1.A~~), (CurUecT2.A)) /
(XprodT1.A + XpodT2.A).

These calculations generate the following joined histogram:

Column T1.A, T2.A

Interval	CurUec	Rows	Value
0	0	0	0
1	1	10,000	25
2	100	200	150

Here we can also calculate row selectivity and UEC selectivity in a similar manner as before:

RowselA = 10,200/80,000 = 0.1275

UecselA = 1/102 = 0.0098.

Comparing the results, we note that approximately 13 times as many rows as the total UEC selectivity would have been produced (i.e. RowselA/UecselA = 13.005). It is this type of skew that the join skew formula corrects when applying multi-column UEC information. If we applied the multi-column formula without correcting for skew we would lose all join skew information.

APPENDIX B

Marked-up Version of the Amended Claims

31. (Amended) A method for optimizing a database management system process of a query, the method comprising:

collecting a plurality of single column statistics for a plurality of columns, the plurality of single column statistics providing estimates for row counts and unique entry counts for a singe column operator;

determining a first selectivity estimate based on an assumption that the columns are substantially independent of each other;

determining a first factor as a measure of a skew of the plurality of columns and as a measure of a dependence of the plurality of **[the]** columns;

determining a second selectivity estimate for predicates in the query using the first selectivity estimate and the first factor, the second selectivity estimate being used in optimizing processing of the query by the database management system.

32. (Amended) The method of claim 31, wherein **the first factor is determined by**

computing a product of unique entry count selectivities **[is calculated]** from **a sum of** maximum unique entry counts for the plurality of columns,

computing a product of maximum initial unique entry counts **[is calculated from maximum initial unique entry counts]** for the plurality of columns,

computing a ratio of the product of unique entry count selectivities and the product of maximum initial entry counts,

selecting a maximum multicolumn unique entry count **[is selected]** from multicolumn entry counts for the plurality of columns, and

computing the first factor **from a** **[is the]** product of **the ratio and an inverse of** **[unique entry count selectivities divided by the product of maximum initial unique entry counts divided by]** the maximum multicolumn unique entry count.